HDF-EOS Aura File Format Guidelines

Cheryl Craig[1], Ken Stone[2], Nathaniel Livesey[3], Steve Friedman[4], Scott Lewicki[4], Doug Ilg[5], Pepijn Veefkind[6], Peter Leonard [7]

August 27, 2001

Version 1.1

1 - National Center for Atmospheric Research
1850 Table Mesa Dr.
Boulder, CO 80303 USA

2 – University of Colorado at Boulder
Center for Limb Atmospheric Sounding
3300 Mitchell Lane, Suite 250
Boulder, CO 80301-2296 USA

3 - Mail Stop 183-701, Jet Propulsion Laboratory
4800 Oak Grove Dr.
Pasadena, California 91109 USA

4 – Mail Stop 169-315, Jet Propulsion Laboratory
4800 Oak Grove Dr.
Pasadena, CA 91109 USA

5 – Raytheon ITSS
4400 Forbes Blvd.
Lanham, MD 20706 USA

6 - KNMI
P.O. Box 201
NL-3730 AE De Bilt The Netherlands

7- Science Systems and Applications, Inc.
10210 Greenbelt Rd, Suite 400
Lanham, MD 20706 USA

NCAR Doc #: SW-NCA-079

HDF-EOS Aura File Format Guidelines

1.0 Introduction

While HDF-EOS constrains HDF with its POINT, SWATH, and GRID implementations, it is still possible to create two files which are completely different and would require dramatically different readers. To ease cross-platform use of Aura data sets, the Aura teams have agreed to make their files match as closely as reasonably possible. The purpose of this paper is to document these guidelines. This standardization only makes sense for Level 2 and above data files and is not intended to apply to Level 1 or other special product files. This document is being written from the standpoint of standardizing Level 2 products, but many of the same points can be used in standardizing Level 3 (and above) as well.

2.0 Areas of potential mismatch

- Major HDF-EOS version (HDF-EOS V2.x and HDF-EOS V5.x are not interchangeable)
- Organization of data fields and attributes
- Dimension Names
- Geolocation Names and dimension ordering
- Data Field names and dimension ordering
- Units for data fields
- Attribute names, values and units

All of the above items can cause difficulties in creating uniform data sets. Each of these items may require a user to write separate code if they differ from other data sets. For this reason we are creating a list of guidelines. It should be understood that each deviation from the set of guidelines will require the user to create special code for each case.

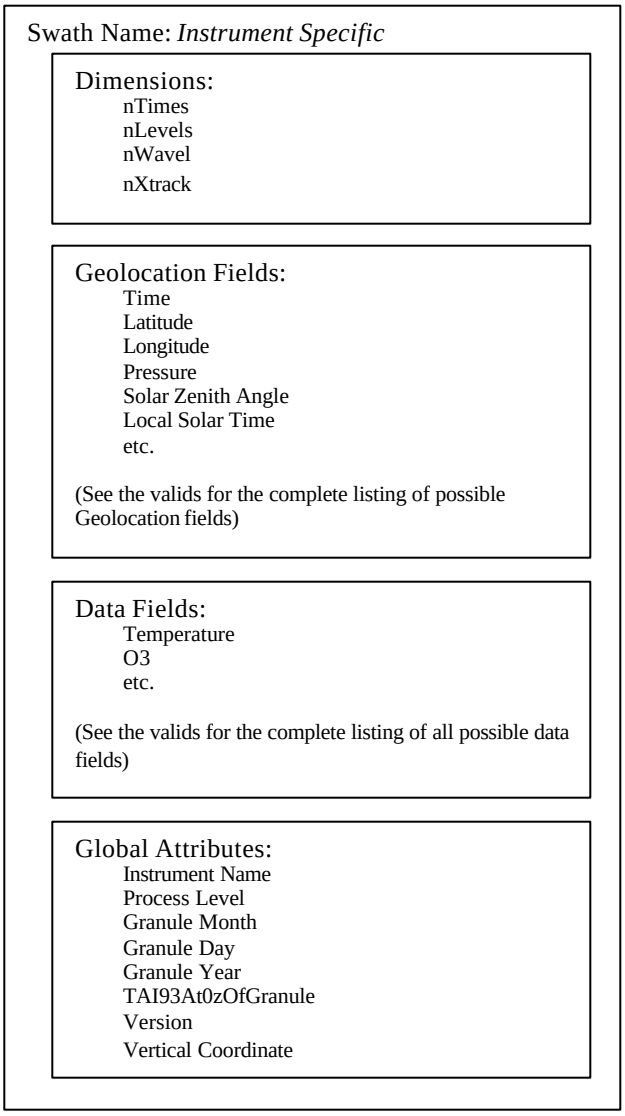
3.0 Guidelines List

- Aura Level 2 files will be created using HDF-EOS V5.x (This guideline is required by launch)
- Aura Level 2 files will use the HDF-EOS Swath format.
- HDF-EOS Structure Names should adhere to the Valids list (in Section 5). It is important to note that in order for names to be uniform, they must match completely. Spacing as well as capitalization must be followed exactly.
- Data should be reported on a pressure grid and be ordered from the ground to space.
- Data Fields should be ordered so that the pressure coordinate is the fastest incrementing coordinate, followed by the number of times (profiles).
- Data Fields should be stored in the units specified.
- HDF fill value and missing data values should have the same value. The actual missing data value will be specified in the MissingValue data field attribute.
- Instrument data fields that are unique to a particular instrument on Aura should be prefixed with the sensor name (e.g., TESBoresightAngle).

4.0 Organization of data fields and attributes

With the adoption of HDF5, the need for two different structures of files has been eliminated. Appendix A describes pre-HDF5 file organization.

More than one species can be contained within a swath and more than one swath can exist within a file.



5.0 Valids

This section lists the specific names which will be used within the file.

5.1 Swath Name

Varies by Instrument

It has become apparent that each instrument has different needs for this name, and standardization is not feasible. A user can use HE5_SWinqswath to determine what the values are for each data set.

5.2 Dimensions:

These are the actual dimensions of the Geolocation and Data field quantities.

nTimes	Number of times (profiles) in data set (this may be unlimited)
nLevels	Number of pressure levels
nWavel	Number of wavelengths
nXtrack	Number of pixels in the across track direction (OMI specific)

5.3 Geolocation Fields:

This is data which describe the scientific measured quantities. They provide information to aid in describing the data's "location".

NOTE THE ORDER OF DIMENSIONS IS IN FORTRAN ORDER WITH THE FIRST DIMENSION BEING THE MOST RAPIDLY INCREMENTING. C USERS SHOULD REVERSE THE DIMENSION ORDER

HDF-EOS Name	*Dimension*	*Suggested units*	*Instrument*				*Notes*
			H	*M*	*O*	*T*	
Time	(nTimes)	seconds from Jan 1, 1993	X	X	X	X	time in TAI units
Latitude	(nTimes)	degrees	X	X		X	Geodetic Latitude
Latitude	(nXtrack,nTimes)	degrees			X		Geodetic Latitude
Longitude	(nTimes)	degrees (-180 to 180)	X	X		X	
Longitude	(nXtrack,nTimes)	degrees (-180 to 180)			X		
Pressure[+]	(nLevels)	hPa (ordered from ground to space)	X	X	X	X	Pressure will be a superset of the UARS pressure levels
Altitude	(nLevels,nTimes)	meters	X			X	
SecondsInDay	(nTimes)	seconds	X	X			Seconds from midnight of day listed in global attributes
SolarZenithAngle	(nTimes)	degrees	X	X	X	X	
LocalSolarTime	(nTimes)	hours	X	X		X	
SpacecraftLatitude	(nTimes)	degrees	X		X	A	Geodetic latitude above WGS84 ellipsoid
SpacecraftLongitude	(nTimes)	degrees (-180 to 180)	X		X	A	Longitude above WGS84 ellipsoid
SpacecraftAltitude	(nTimes)	meters	X		X	A	Height above WGS84 ellipsoid
OrbitAscendingFlag	(nTimes)	logical				A	if true then orbit is ascending
OrbitGeodeticAngle	(nTimes)	degrees		X			
LineOfSightAngle	(nTimes)	degrees (East of North)		X			
SolarAzimuthAngle	(nXtrack,nTimes)	degrees (East of North)			X	A	
ViewingZenithAngle	(nXtrack,nTimes)	degrees			X		
ViewingAzimuthAngle	(nXtrack,nTimes)	degrees (East of North)			X		
SpacecraftAzimuth	(nTimes)	degrees (East of North)				A	
SpacecraftZenith	(nTimes)	degrees				A	
HIRDLSScienceScanMode	(nTimes)	short integer	X				HIRDLS Science Scan Mode identifier
HIRDLSScanUpFlag	(nTimes)	logical	X				HIRDLS Scan Up identifier

HDF-EOS Name	*Dimension*	*Suggested units*	*Instrument*				*Notes*
			H	*M*	*O*	*T*	
HIRDLSScanElevationAtNominalAltitude	(nTimes)	degrees	X				
HIRDLSScanAzimuthAtNominalAltitude	(nTimes)	degrees	X				
HIRDLSTangentHeightAtNominalAltitude	(nTimes)	meters	X				
HIRDLSViewDirectionAtNominalAltitude	(nTimes)	degrees (East of North)	X				
TESDate	(nTimes)					A	TES survey may cover multiple days
TESSequenceID	(nTimes)	sequence within current survey				X	
TESBoresightNadirAngle	(nTimes)	degrees				A	32 bits (accommodate N+N2 scans)
TESFiltersUsed	(nTimes)	bit mask				A	
TESPixelsUsedFlag	(nTimes)	bit mask				A	
TESRetrievedPointingAngle	(nTimes)	degrees				A	(Limb Only)
TESSurfaceAltitude	(nTimes)	meters				A	Altitude of surface or lowest level retrieved in scenes with clouds

X – Field in standard file

A – Data items which will be carried in an ancillary file that is created only once per set of standard products

+ nLevels for TES is currently set to 14 levels (based on UARS standard) but may be changed to 85 levels.
*nLevels for HIRDLS is set to 145 (1000. * 10 ** (-i/24) i=0,144)*

5.4 Data Field Names:

This is the actual scientific data.

NOTE THE ORDER OF DIMENSIONS IS IN FORTRAN ORDER WITH THE FIRST DIMENSION BEING THE MOST RAPIDLY INCREMENTING. C USERS SHOULD REVERSE THE DIMENSION ORDER

HDF-EOS Name	*Dimension*	*Suggested units*	*H*	*M*	*O*	*T*	*Notes*
Temperature	(nLevels,nTimes)	K	X	X		X	
TemperaturePrecision	(nLevels,nTimes)	K	X	X		X	
TemperatureNormChiSq	(nLevels,nTimes)	K	X				
O3	(nLevels,nTimes)	vmr	X	X		X	
O3Precision	(nLevels,nTimes)	vmr	X	X		X	
O3NormChiSq	(nTimes)		X				
O3	(nLevels,nXtrack,nTimes)	vmr			X		OMI imager data
H2O	(nLevels,nTimes)	vmr	X	X		X	
H2OPrecision	(nLevels,nTimes)	vmr	X	X		X	
H2ONormChiSq	(nTimes)		X				
ClONO2	(nLevels,nTimes)	vmr	X				
ClONO2Precision	(nLevels,nTimes)	vmr	X				
ClONO2NormChiSq	(nTimes)		X				
N2O5	(nLevels,nTimes)	vmr	X				
N2O5Precision	(nLevels,nTimes)	vmr	X				
N2O5NormChiSq	(nTimes)		X				
N2O	(nLevels,nTimes)	vmr	X	X			
N2OPrecision	(nLevels,nTimes)	vmr	X	X			
N2ONormChiSq	(nTimes)		X				
NO2	(nLevels,nTimes)	vmr	X			X	TES limb only
NO2Precision	(nLevels,nTimes)	vmr	X			X	TES limb only
NO2NormChiSq	(nTimes)		X				
CH4	(nLevels,nTimes)	vmr	X			X	
CH4Precision	(nLevels,nTimes)	vmr	X			X	
CH4NormChiSq	(nTimes)		X				
HNO3	(nLevels,nTimes)	vmr	X	X		X	TES limb only
HNO3Precision	(nLevels,nTimes)	vmr	X	X		X	TES limb only
HNO3NormChiSq	(nTimes)		X				

HDF-EOS Name	*Dimension*	*Suggested units*	*H*	*M*	*O*	*T*	*Notes*
CFC11	(nLevels,nTimes)	vmr	X				
CFC11Precision	(nLevels,nTimes)	vmr	X				
CFC11NormChiSq	(nTimes)		X				
CFC12	(nLevels,nTimes)	vmr	X				
CFC12Precision	(nLevels,nTimes)	vmr	X				
CFC12NormChiSq	(nTimes)		X				
OH	(nLevels,nTimes)	vmr		X			
OHPrecision	(nLevels,nTimes)	vmr		X			
HO2	(nLevels,nTimes)	vmr		X			
HO2Precision	(nLevels ,nTimes)	vmr		X			
CO	(nLevels,nTimes)	vmr		X		X	
COPrecision	(nLevels,nTimes)	vmr		X		X	
HCN	(nLevels,nTimes)	vmr		X			
HCNPrecision	(nLevels,nTimes)	vmr		X			
NO	(nLevels,nTimes)	vmr				X	TES limb only
NOPrecision	(nLevels,nTimes)	vmr				X	TES limb only
HCl	(nLevels,nTimes)	vmr		X			
HClPrecision	(nLevels,nTimes)	vmr		X			
HOCl	(nLevels,nTimes)	vmr		X			
HOClPrecision	(nLevels,nTimes)	vmr		X			
ClO	(nLevels,nTimes)	vmr		X			
ClOPrecision	(nLevels,nTimes)	vmr		X			
BrO	(nLevels,nTimes)	vmr		X			
BrOPrecision	(nLevels,nTimes)	vmr		X			
CirrusIceContent	(nLevels,nTimes)	g/m^3		X			
CloudTopHeight	(nTimes)	meters				X	
CloudTopPressure	(nXtrack,nTimes)	hPa			X		(OMI imager data)
CloudTopPressure	(nTimes)	hPa	X			X	
ColumnAmountO3	(nXtrack,nTimes)	Dobson units		X	X		
TropoColumnAmountO3	(nXtrack,nTimes)	Dobson units			X		
ColumnAmountNO2	(nXtrack,nTimes)	$molec/cm^2$			X		
ColumnAmountSO2	(nXtrack,nTimes)	$molec/cm^2$			X		
ColumnAmountBrO	(nXtrack,nTimes)	$molec/cm^2$			X		
ColumnAmountHCHO	(nXtrack,nTimes)	$molec/cm^2$			X		

HDF-EOS Name	*Dimension*	*Suggested units*	*H*	*M*	*O*	*T*	*Notes*
ColumnAmountOClO	(nXtrack,nTimes)	molec/cm^2			X		
CloudFraction	(nXtrack,nTimes)	dimensionless			X		
AerosolOpticalThickness	(nWavel,nXtrack,nTimes)	dimensionless			X		
AerosolSingleScatteringAlbedo	(nWavel,nXtrack,nTimes)	dimensionless			X		
UVAerosolIndex	(nXtrack,nTimes)	dimensionless			X		
VisibleAerosolIndex	(nXtrack,nTimes)	dimensionless			X		
SurfaceTemperature	(nTimes)	K				X	(May be cloud top for limb or may be omitted for limb)
RandomError	(nLevels,nTimes)	range: 0-1				X	
LandSurfaceEmissivity	(nTimes)	range: 0-1				A	
SystematicError	(nLevels,nTimes)	vmr or temp				X	Vmr for species, temp for temp retrieval
TotalColumnDensity	(nTimes)	moles/cm^2				X	TES may still split into Trop/Strat
TotalColumnDensityError	(nTimes)	moles/cm^2				X	TES may still split into Trop/Strat
EstRandomErrors	(nLevels,nTimes)	range: 0-1				X	
EstSystematicErrors	(nLevels,nTimes)	range: 0-1				X	
FractionalExplainedVariance	(nLevels,nTimes)	range: 0-1				X	
CorrelationLength	(nLevels,nTimes)	0-1000 km				X	
Convergence	(nTimes)					X	(Chi^2) final retrieval – *not normalized*
CompConvergenceToEstimate	(nTimes)					A	Comparison between actual Chi^2 and estimated Chi^2 – *not normalized*
FinalRetrievalConverged	(nTimes)	0-1 (Boolean)				A	Final retrieval converged, 1=T=successful
FinalRetrievalnIterations	(nTimes)	0-8 (small int.)				A	Number of model iterations for final retrieval
SpeciesRetrivalConverged	(nTimes)	0-1 (Boolean)				X	Species retrieval converged, 1=T=success
SpeciesRetrievalnIterations	(nTimes)	0-8 (small int.)				X	Number of model iterations for retrieval the particular species
OutOfExpectedRange	(nTimes)	0-1 (Boolean)				X	1=T=retrieved results out of expected range/norm
RMSResidual	(nTimes)					X	
MaxRMSResidual	(nTimes)					X	
CloudCover	(nTimes)	0-1 (Boolean)				A	Identified through retrieval or L1B, 1=T=suspected cloud cover identified
ProcessingTime	(nTimes)					A	Time to process full retrieval
DegreesOfFreedomFromSignal	(nTimes)	0-100				X	Max value <= number of levels (87). Indicates information content vs. signal.

HDF-EOS Name	*Dimension*	*Suggested units*	*H*	*M*	*O*	*T*	*Notes*
JacobianMethod	(nTimes)	text (“calc” or “lookup”)				X	
UVSurfaceIrradiance	(4,nXtrack,nTimes)	Watts/m^2/nm			X		Spectral surface irradiances at 305, 310, 324 and TBD nm
UVSurfaceIrradianceEW	(nXtrack,nTimes)	Watts/m^2			X		Erythemally weighted UV surface irradiance

X – Field in standard file

A – Data items which will be carried in an ancillary file that is created only once per set of standard products

6.0 Definition of Possible Attributes in File

The following attributes are ones which may appear in the file. They are meant to provide additional information or to ease use of the data. For instance, while the date is provided in the ECS attached metadata, the GranuleMonth/Day/Year are provided as a simpler interface to this information.

6.1 File Level Attributes:

This is information which helps to describe this particular data set. It can be useful in labeling plots, calculating dates, etc.

Attribute Name	*Attribute Type*	*Attribute Description*
InstrumentName	HDFE_NATIVE_CHAR	"HIRDLS", "MLS", "TES" or "OMI"
ProcessLevel	HDFE_NATIVE_CHAR	Processing Level --- "2", " 3" etc.
GranuleMonth	HDFE_NATIVE_INT	Month of start of granule --- 1-12
GranuleDay	HDFE_NATIVE_INT	Day of start of granule ---- 1-31
GranuleYear	HDFE_NATIVE_INT	Year of start of granule ---- i.e. 2003
TAI93At0zOfGranule	HDFE_NATIVE_INT	TAI time of 0z of granule
Version	HDFE_NATIVE_CHAR	Processing version

6.2 Swath Level Attributes

This is information which helps describe the swath to which it is attached.

Attribute Name	*Attribute Type*	*Attribute Description*
Pressure*	HDFE_NATIVE_REAL	**MANDATORY** – pressure levels
VerticalCoordinate	HDFE_NATIVE_CHAR	"Pressure", "Altitude", "Potential Temperature"

* This attribute is an exact duplicate of the Pressure Geolocation Field. Writing the pressure data in two locations was agreed upon as a compromise between instrument teams.

6.3 Data Field Attributes

This is information which helps to describe the individual data fields. Data Field Attributes are a feature which can be useful in annotating plots as well as describing the data product to input routines. If ScaleFactor and Offset are not applicable, a value of 1 and 0 may be supplied respectively.

Attribute Name	*Attribute type*	*Attribute Description*
MissingValue	Same type as Data Field	Contains the value for missing data
ScaleFactor	HDFE_NATIVE_FLOAT	Factor for scaling data (if applicable)
Offset	HDFE_NATIVE_FLOAT	Value to add to the data (if applicable)
Title	HDFE_NATIVE_CHAR	For labeling a plot or axis
Units	HDFE_NATIVE_CHAR	Labeling units (for labeling color bars, converting between units, etc). After applying scale and offset.

Values for Units: "m", "km", "hPa", "Pa", "mAtm-cm", "vmr", "K", "micron", "degree", "second", "hour", "m-1", "dimensionless", etc.

Appendix A
File Formats before HDF-EOS5

It is assumed that each individual instrument team will transition to HDF-EOS5 based on their own time schedules. There are several items which will most likely be different in these pre-HDF-EOS5 data sets.

Local data attributes are cumbersome to implement and use in HDF-EOS2, so implementation of these features before HDF-EOS5 are optional.

Two types of files exist until HDF-EOS5 is used.

One or more species per file
Type 1

Swath Name: *Instrument Specific*

Dimensions:
- nTimes
- nLevels

Geolocation Fields:
- Time
- Latitude
- Longitude
- Pressure
- Solar Zenith Angle
- Local Solar Time

(See the valids for the complete listing of possible Geolocation fields)

Data Fields:
- Temperature
- O3
- etc.

(See the valids for the complete listing of all possible data fields)

Global Attributes:
- Instrument Name
- Process Level
- Granule Month
- Granule Day
- Granule Year
- Version

Only one species per file
Type 2

Swath Name: *Measured Field Name*

Dimensions:
- nTimes
- nLevels

Geolocation Fields:
- Time
- Latitude
- Longitude
- Pressure
- Solar Zenith Angle
- Local Solar Time

(See the valids for the complete listing of possible Geolocation fields)

Data Fields:
- L2gpValue
- L2gpPrecision

Global Attributes:
- Instrument Name
- Process Level
- Granule Month
- Granule Day
- Granule Year
- Version

Type 1:
Swath Name is instrument specific (there is no guideline for this)
Data field names are the species names
This allows for the possibility of more than one species to be contained within a file and more than one swath per file as well.

Type 2:
Swath Name is the species name
L2gpValue is always the name of the data field
This makes it easier to write a generic reader.

When HDF5 is adopted, the Type 2 will no longer be needed. HDF-EOS5 has the ability to create links (similar to symbolic links in UNIX) and the name L2gpValue will be a link while the data field name will be from the valids list.